# Folding of small proteins: A matter of geometry?

PFN Faísca[1,*] and MM Telo da Gama[1]

[1] *Centro de Fisica Teórica e Computacional da Universidade de Lisboa, Av. Prof. Gama Pinto 2, 1649-003 Lisboa Codex, Portugal*

**We review some of our recent results obtained within the scope of simple lattice models and Monte Carlo simulations that illustrate the role of native geometry in the folding kinetics of two state folders.**

PACS numbers: **87.15.Cc; 91.45.Ty**
Keywords: **kinetics, funnels, contact order, cooperativity, long-range contacts**

## I. INTRODUCTION

During the 1960s the Nobel Laureate Christian Anfinsen showed through *in vitro* experiments that globular proteins are capable of spontaneously self-assemble into their complex, three-dimensional native (i.e., biologically active) structures through the process of protein folding [1]. This finding suggested that the only information required for a given protein sequence to fold into its unique native state is the sequence itself, an idea that gave rise to the so-called "protein folding" problem: The prediction of the native fold from the knowledge of the protein sequence. This question which, in more fundamental terms, concerns the understanding of the physical principles involved in the mechanisms of folding and how the latter, together with the observed folding rates, are encoded in the protein's amino acid sequence has dominated the folding literature up to the late 1990s.

In the late 1960s Cyrus Levinthal raised a problem related to Anfinsen's observation that the native state is the global minimimum of the free energy [2]. By using a simple counting argument, Levinthal quantified the magnitude of the folding process and concluded that proteins would not fold in a reasonable amount of time if the search for the native conformation is performed randomly as may be inferred from Anfinsen's observation, an idea that became known as the Levinthal paradox.

An important contribution to protein folding research, that proposes a new scenario for folding, is that of the energy landscape theory (ELT), put forward by Bryngelson and Wolynes in the late 1980s [3]. These authors recognised that akin to spin glasses proteins exhibit frustration. Indeed, two kinds of frustration may be distinguished in proteins: energetic frustration, owing to unfavourable interactions present in the native state, and geometric frustration that leads to energy barriers between structurally related conformations and results from geometric constraints (e.g., chain connectivity). A possible scenario is an energy landscape (i.e., the free energy as a function of one or more conformation coordinates) similar to that of random heteropolymers (RHP), with many local minima separated by energy barriers and where high energy barriers are likely to produce kinetic traps, that is, long-lived low free energy conformations. However, analytical studies of RHPs together with Monte Carlo simulations provided evidence that, by contrast with typical RHPs, the energy landscapes of kinetically foldable lattice proteins are smooth, with a large free energy gap separating the native state from misfolded conformations [4], and its overall slope is such that the protein is easily driven down to the native state (i.e., it must be funnel shaped) [5]. Lattice proteins are models that reduce the protein backbone to a string of single site beads. While it is true that such minimalist models do not capture the full complexity of real proteins they are non-trivial models that describe some potentially relevant aspects of protein folding kinetics. See [6] for a recent review on computational methods in protein folding that discusses the strengths and limitations of lattice models.

In addition to providing deep new insights into the protein folding process the 'energy landscape perspective' has stimulated an invaluable synergy between theoretical and experimental research in the field of protein folding. In particular, it allowed for "new interpretations of existing experiments and has led to the design of new strategies to probe the details of the folding process" [10]. A clear example of successful theoretical prediction that shaped modern thinking about protein folding kinetics is that of the concept of nucleation-growth (or -condensation) mechanism. The latter was firstly discovered by Shakhnovich and his collaborators in the context of a lattice model and Monte Carlo folding simulations [7] and confirmed by Fersht [8] using a protein engeneering method termed $\phi$-value analysis [9]. Furthermore, the above mentioned criteria for kinetically foldable lattice-polymers have been tested for real proteins. The experimental results suggest that, in spite of their success in predicting biologically relevant time frames for folding, smooth energy landscapes and large energy gaps do not account for the remarkable six order of magnitude range characteristic of the folding rates of real two-state folders [11]. By contrast, a parameter of native geometry, that measures the average sequence separation of contacting residue pairs in the native fold, named contact order (CO), appears to be strongly coupled to the

kinetics of many small (with ≈ up to 120 amino acids) protein molecules exhibiting two-state folding kinetics. Indeed, Plaxco *et al.* [12, 13] found a strong correlation ($r = 0.92$) between the CO and the folding rates of 24 single-domain, two-state folders. More recently, other measures of the native geometry were proposed and were found to correlate as well as the CO with real two-state folding kinetics [14, 15], supporting the idea that the physics underlying the folding mechanism of these folders, although 'encoded' in the protein's primary sequence, may not depend strongly on its finer details. These observations triggered a renewed interest in the effects of the native state's geometrical properties on protein folding, an issue originally addressed by Gō and Taketomi in a pioneering study based on a two-dimensional protein lattice model where the relative role of local and non-local contacts in the energetics and kinetics of folding [16] were investigated.

In this paper we review some recent results obtained within the scope of lattice models and Monte Carlo simulations that address the role of native geometry in determining the folding kinetics of small lattice proteins.

## II. LATTICE MODELS AND MONTE CARLO SIMULATIONS

The lattice-polymer model discretizes space by embedding the protein in a regular three-dimensional lattice. The protein is reduced to its backbone structure: amino acids are represented by beads of uniform size, occupying the lattice vertices and the peptide bond, that covalently connects amino acids along the polypeptide chain, is represented by sticks, with uniform length, corresponding to the lattice spacing ( Figure 1). In order to satisfy the excluded volume constraint only one bead is allowed per lattice site.

### The Gō model

In the Gō model the energy of a conformation, defined by the set of bead coordinates $\{\vec{r_i}\}$, is given by the contact Hamiltonian

$$H(\{\vec{r_i}\}) = \sum_{i>j}^{N} \epsilon \Delta(\vec{r_i} - \vec{r_j}), \qquad (1)$$

where the contact function $\Delta(\vec{r_i} - \vec{r_j})$, is unity only if beads $i$ and $j$ form a non-covalent native contact, i.e., a contact between a pair of beads that is present in the native structure, and is zero otherwise. The Gō potential is based on the idea that the native fold is very well optimised energetically. Accordingly, it ascribes equal stabilizing energies ($\epsilon < 0$) to all the native contacts and neutral energies ($\epsilon = 0$) to all non-native contacts.

### The Shakhnovich model

By contrast with the Gō model, which ignores the protein's chemical composition, the Shakhnovich model [17] addresses the dependence of protein folding dynamics on the amino acid sequence by considering interactions between the 20 different amino acids used by Nature in the synthesis of real proteins. Accordingly, the contact Hamiltonian that defines the energy of each conformation is such that

$$H(\{\sigma_i\}, \{\vec{r_i}\}) = \sum_{i>j}^{N} \epsilon(\sigma_i, \sigma_j) \Delta(\vec{r_i} - \vec{r_j}), \qquad (2)$$

where $\{\sigma_i\}$ represents an amino acid sequence, and $\sigma_i$ stands for the chemical identity of bead $i$. In this case both native and non-native contacts contribute energetically to the folding process. Therefore, the contact function $\Delta$ is 1 if any two beads $i$ and $j$ are in contact but not covalently linked and is 0 otherwise. The interaction parameters $\epsilon$ are taken from the $20 \times 20$ Miyazawa-Jernigan (MJ), derived from the distribution of contacts of real native proteins [18].

### Simulation details

To mimic protein motion in simple lattice models, a Monte Carlo (MC) algorithm is used together with the kink-jump move set. This means that local random displacements of one or two beads are repeatedly accepted or rejected in accordance with the Metropolis rule [19]. A MC run starts from a randomly generated unfolded conformation (typically with very few native contacts) and the folding dynamics is traced by following the evolution of the fraction of native contacts, $Q = q/Q_{max}$, where $Q_{max}$ is the total number of native contacts for each chain length, and $q$ is the number of native contacts at each MC step. Kinetic quantities such as the folding time, $t$, is taken as the first passage time (FPT), that is, the number of MC steps that corresponds to $Q = 1.0$. All native structures considered are maximally compact cuboids found by homopolymer relaxation [36].

The folding dynamics is studied at the so-called optimal folding temperature, the temperature that minimizes the folding time as measured by the mean FPT [21, 22, 23, 24].

The sequences studied within the context of the Shakhnovich model were prepared using the design method developed by Shakhnovich and Gutin (SG) [20] based on random heteropolymer theory and simulated annealing techniques.

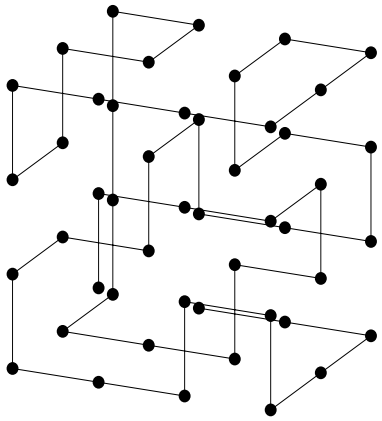

FIG. 1: A maximally compact conformation (MCC), representing the native structure, for a 48 bead long chain in a cubic lattice. MCCs, with a maximal number of contacts between amino acids for each chain length, mimic the high degree of compactness that characterises real protein native structures.

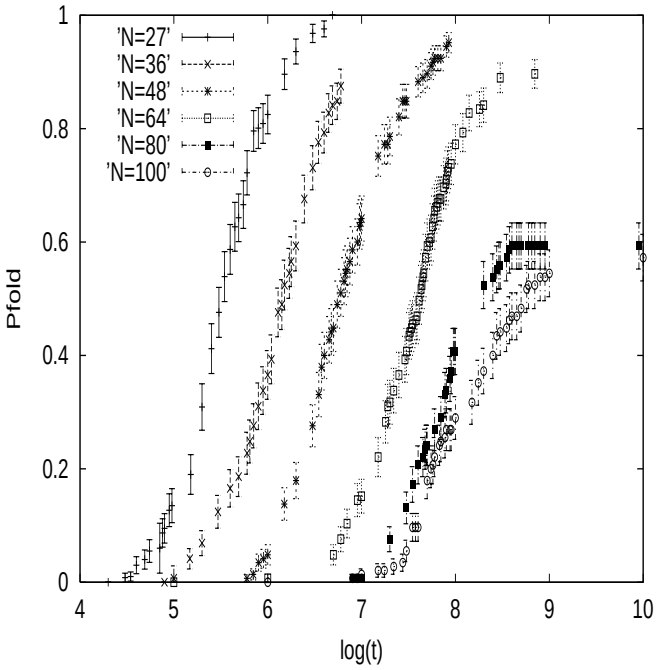


FIG. 2: Dependence of the folding probability, $P_{fold}$, on $log_e(t)$. $P_{fold}$ was computed as the fraction of folding simulations that ended up to time $t$ normalized to the total number of MC runs performed for each chain length. For $N$ up to 64 the curves are consistent with asymptotic values of $P_{fold} \to 1$. For $N \geq 80$ this dynamical behaviour breaks down and the asymptotic value of $P_{fold}$ decreases considerably [24].

## III. CHAIN LENGTH AS A DETERMINANT OF FOLDING KINETICS

There is empirical evidence that the protein's chain length, $N$, grossly determines the kinetics of folding according to the following rule: for $N$ up to $\approx 100$ amino acids the kinetics is two-state, while for larger $N$ protein chains typically exhibit non-two-state kinetic behaviour [25] (there are of course exceptions to this rule and it is possible to find proteins with chain lenght $N \leq 110$ that fold via a three-state kinetics [26]). When the kinetics is two-state, folding proceeds in the absence of any observable intermediates and there is a single transition state associated with one major free energy barrier separating the native from the unfolded conformations [26].

What do lattice models tell us about the role of chain length in the protein folding kinetics? Results obtained by one of us [24] in the context of the Shakhnovich lattice-polymer model showed that, depending on the chain length, two dynamical regimes for folding can be identified (Figure 2). In the regime found for $N \geq 80$ the folding performance depends on the native state's structure, with certain structures being kinetically more accessible (i.e., more easily foldable) than others. This observation is in agreement with results reported in Ref. [27] where it is found that, for 125 bead long proteinlike sequences, efficient folding depends on structural features of the native state (e.g., the distribution and position of contacts in the native structure). While it is likely that kinetic traps dominate folding in this regime (akin to what happens in the folding of real longer proteins) we have not conducted our MC simulations long enough to conclude that this is actually the case.

For $N < 80$, as in real two-state proteins, we have found that kinetic relaxation is well described by a single exponential law (a distinguishing feature of a two-state kinetic process) with the reactant concentration (the equivalent in our simulations to the fraction of unfolded chains) being proportional to $\exp^{-kt}$ where $k$ is the so-called relaxation rate constant (Fig. 3).

As for the dependence of folding time on the chain length we have found a scaling law of the type $t \sim N^5$ [24] while, for the same model, Gutin *et al.* [21] have reported that $t$ and $N$ scale as $t \sim N^4$. For the Gō model, on the other hand, a weaker dependence of $t \sim N^3$ has been observed [21, 22]. These findings are in broad agreement with Thirumalai's theoretical prediction that, near the point of thermodynamic equilibrium between the unfolded coil and the native fold, i.e., at the folding transition temperature, the folding time and protein size scale as $t \sim N^\lambda$ with $\lambda$ between 3.8 and 4.2 [28]. For real proteins, however, it has been shown that the dependence

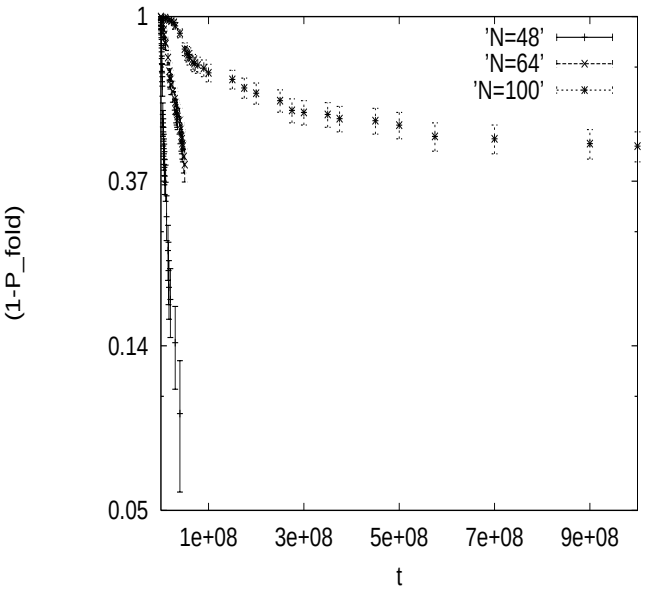


FIG. 3: Evidence for single exponential kinetic relaxation for $N = 48$ and $N = 64$ but not for $N = 100$. The correlation coefficient between the logarithmic fraction of unfolded chains (i.e., the 'reactant concentration') and the folding time $t$ is $r = 0.99$ for $N = 48$ and $N = 64$ [24].

of the folding time on protein chain length is weak (r = 0.16) [12], and recent results by Galzitskaya *et al* [29] suggest that protein size is the main determinant of folding rates for multi-state proteins only. We note that lattice simulations use single or double residue MC move sets which may differ very significantly from the collective motions of large portions of the chain that can occur in continuum space. Thus the dependence of the folding times of lattice polymers on the chain length could differ significantly from that observed in real proteins. This shortcoming of the model may be particularly relevant for large protein molecules.

## IV. NATIVE GEOMETRY AS A DETERMINANT OF FOLDING RATES

In 1998 Plaxco *et al.* [12] proposed the relative contact order parameter, CO, as a simple empirical measure of the native structure's geometric complexity. The CO measures the average sequence separation of contacting residue pairs in the native structure relative to the chain length of the protein and is defined as

$$CO = \frac{1}{LN} \sum_{i,j}^{N} \Delta_{i,j} |i - j|, \qquad (3)$$

where $\Delta_{i,j} = 1$ if residues $i$ and $j$ are in contact and is 0 otherwise; $N$ is the total number of contacts and $L$ is the protein chain length. In a subsequent study Plaxco and co-workers reported a rather strong correlation ($r = 0.92$) between the CO parameter and the logarithmic folding rates of 24 single domain, two-state folders [13], suggesting that the native's state geometry could be the major determinant of two-state folding kinetics.

Do lattice models reproduce the experimentally observed correlation between folding rates and the CO? We have recently addressed this question in the context of the Shakhnovich model. A set of 20 target native structures, selected in order to cover the range of CO observed in real proteins, was investigated for chain lengths $N = 36, 48, 54, 64$ and 80. A correlation of $r = 0.70 - 0.79$ was found between increasing CO and longer logarithmic folding times for chain lengths $N \geq 54$ [36]. There is a potential shortcoming of using the MJ potential to investigate contact order correlations with folding rates in lattice models. Indeed, it was shown recently that the energy landscapes of real two-state folders are relatively smooth [37, 38] while the folding dynamics associated with the MJ potential is prone to energetic traps (that result from competing interactions between pairs of beads) that may lead to considerably rough energy landscapes. Moreover, if the geometry-related kinetics of real two-state folders is favoured by their smooth energy landscapes and if the relevant topological effects are subtle they could be masked by stronger kinetic effects arising in sequence-specific models such as the Shakhnovich [39]. Following this line of reasoning Jewett *et al.* [39] studied the correlation of the folding time on the contact order parameter in the context of the Gō model. Since the Gō potential considers only attractive interactions between native contacts the corresponding folding dynamics is prone to geometric traps only, i.e., traps that result from chain connectivity and the geometry of the native fold. Thus, as real two-state proteins, lattice-polymers modelled by Gō and Gō-type interaction schemes exhibit relatively smooth energy landscapes making them particularly suitable models to investigate the role of the native state's geometry in the folding kinetics. Notwithstanding, Jewett *et al.* found a very poor correlation, $r \approx 0.23$, between logarithmic folding times and the CO in a pool of targets comprising 97 different native structures with chain length $N = 27$. A modified version of the original Gō model, characterized by a non-linear relationship between the native state's energy and the number of native contacts formed during folding, i.e., that enhances folding cooperativity, exhibits a stronger correlation (r=0.75) between the logarithmic folding time and the contact order [39]. Although this correlation is weaker than the correlation observed experimentally (and lower than that found by us for the Shakhnovich model) this finding suggests that the physical mechanism underlying geometry-dependent kinetics may be that of cooperativity.

In a recent study we have identified different folding mechanisms, distinguished by different cooperativities, for the Shakhnovich model [40]. For low contact order

structures we observed that the building up of the native fold may occur in a gradual manner, where the amount of native structure increases in a continuous way as time evolves, or in a more abrupt (or cooperative) way where a significant portion of native structure emerges only during the late stages of folding. By contrast, the folding of intermediate and high contact order structures is clearly more cooperative, in the sense described above. Indeed, in the latter case, it is possible to identify a clear pattern in the formation of the native fold that is driven by the backbone distance: local contacts (i.e. close in space and in sequence) form first and long-range contacts (i.e. close in space but distant in sequence) form progressively later as contact range increases. As a consequence a monotonic decrease of contact frequency with increasing contact range is observed and this trend is specific of the folding dynamics of low contact order structures. These results provide a possible explanation of the higher correlation found in Ref. [36] between contact order and logarithmic folding times for chain length $N \geq 54$ and predominantly high-CO values. If cooperativity is the essential ingredient of geometry-dependent kinetics, as the results on modified Gō models suggest, and if the long-range (LR) contacts enhance the cooperativity of the folding transition it is natural to expect a stronger correlation in high contact order structures, which have predominantly LR contacts.

## V. THE ROLE OF LOCAL AND LONG-RANGE CONTACTS: REVISITING THE GŌ MODEL

Contact order is a way to quantify a protein's geometry that accounts for the average range of amino acid interactions in the native fold. It is interesting to note that in one of the very first studies that made use of a lattice-polymer framework to model folding, Gō and Taketomi investigated the role played by local and long-range (LR) inter-residue interactions in the dynamics of this complex biological process [16]. The latter has actually became one of the most widely-debated issues in the protein folding literature. As a result of this debate there is now agreement on the fact that LR interactions play an important role in stabilizing the native fold [16, 30, 31, 32] but there is no consensus on their role in the folding kinetics. For example, early results of Gō and Taketomi [16] for a 49-residue chain on a two-dimensional square lattice suggest that local interactions accelerate both the folding and unfolding transitions. In Ref.[33] Unger and Moult have studied optimised heteropolymer sequences with chain length $N = 27$ on a three-dimensional cubic lattice and concluded that increasing the strength of local interactions increases the ability of sequences to fold. By contrast, results obtained by Abkevich *et al.*[30] for the Shakhnovich lattice-polymer model provided evidence that, under conditions where the native state is stable, a 36-residue sequence on a three-dimensional cubic lattice folds to a native structure with mostly LR contacts two-orders of magnitude faster than a sequence folding to a native structure with predominantly local contacts. In Ref. [34] Govindarajan and Goldstein have used a lattice model in conjunction with techniques drawn from the theory of spin glasses and found that optimal conditions for folding are achieved when local interactions contribute little to the native state's energy.

The finding that the CO is correlated with the folding kinetics of small, two-state proteins has set a new ground for investigating the role of local (and LR) contacts in protein folding, one where the effects of native geometry are taken into account. Moreover, it is natural to address this issue in the context of the Gō model since lattice polymers modelled by the Gō potential exhibit smooth free energy landscapes where the effects of native geometry dominate. We have revisited the Gō model in the light of these findings to investigate the role of local (and LR) inter-residue interactions in the dynamics of folding [35]. We introduced a parameter $\sigma$ that weights the relative contributions of local and long-range interactions to the total energy of the native state. When $\sigma = 0$ all LR native interactions are 'switched-off' and only local interactions contribute to the total energy of a conformation. The opposite situation is observed when $\sigma = 1$. We considered two energy parametrizations, one at fixed native state's energy, and three target structures, with different geometries. We found that, when the native state's energy varies with $\sigma$, the native fold always exhibits the highest occupation probability, a measure of its stability, and that, with the exception of one structure, the latter lies between 0.6 and 0.9 [35].

Our results show that LR interactions play a major role in determining the folding kinetics of 48-mer three-dimensional lattice polymers modelled by the Gō potential. Indeed, for three target structures, with different native geometries, we observed a sharp increase in the folding time when the relative contribution of the LR interactions to the native state's energy is decreased. However, the kinetic response to a decrease in the relative contribution of the LR interactions is strongly dependent on the target geometry. In fact, we have observed a remarkable three-order of magnitude span in the folding time of Gō polymers folding to one of the target structures studied (Figure 4).

The existence of different kinetic responses as a function of target geometry has a mechanistic interpretation. We have found that, for a given target geometry, a geometric coupling exists between local and LR contacts. When this is the case, the establishment of LR contacts is forced by the (previous) formation of local contacts. The absence of this geometric coupling leads to kinetics that are sensitive to the interaction energy parameters; in this case establishment of the local contacts is not sufficient to promote the formation of the LR ones if they

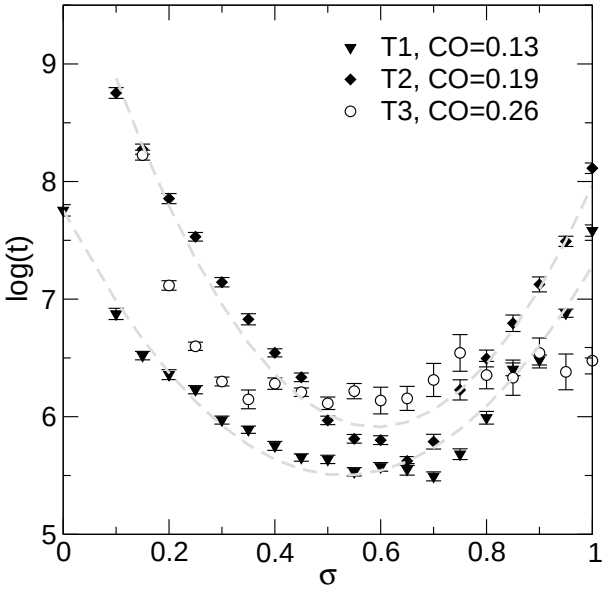


FIG. 4: Dependence of the logarithmic folding time, $\log_{10}(t)$, on the long-range interaction parameter, $0 \leq \sigma \leq 1$ for three target geometries at fixed native state's energy. The fraction of LR contacts is 0.77 for target $T3$, while for targets $T1$ and $T2$ it is 0.48 and 0.42, respectively. For the three structures the folding time increases considerably faster when $\sigma$ decreases than when $\sigma$ increases away from the respective minima. The behaviour observed for $T3$ is trivial and results from its considerably high content in LR contacts. In the limit of $\sigma = 0$, the structure is forced to fold with only 20 per cent of its native interactions and this results in folding failure. The results obtained for the low- and intermediate-CO target structures, $T1$ and $T2$, are more interesting. The corresponding curves are qualitatively similar but a closer inspection reveals an important difference, namely: for $\sigma < 0.5$ the dependence of the folding time on $\sigma$ is much stronger for the intermediate-CO structure, $T2$. Indeed, in this case one observes a remarkable three-order of magnitude dispersion of folding times, ranging from $\log_{10}(t_{min}) = 5.76 \pm 0.05$ (for $\sigma = 0.65$) to $\log_{10}(t_{max}) = 8.75 \pm 0.05$ (for $\sigma = 0.10$), by contrast with $T1$ for which $\log_{10}(t_{min}) = 5.50 \pm 0.08$ (for $\sigma = 0.70$) and $\log_{10}(t_{max}) = 7.69 \pm 0.09$ (for $\sigma = 0.00$) [35].

are strongly penalized energetically, resulting in longer folding times.

## VI. PROTEINLIKE COOPERATIVITY: A NEW CHALLENGE FOR LATTICE MODELS

A well-established criterion for two-state thermodynamic cooperativity observed in folding experiments of real proteins is the calorimetric criterion introduced in

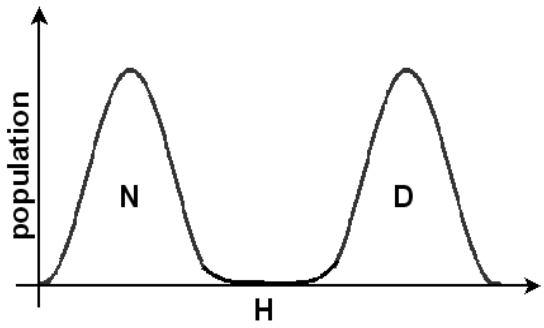


FIG. 5: Calorimetric criterium: the distribution of enthalpy at the midpoint of a two-state folding transition is bimodal with very few molecules having enthalpies between the native (N) and denatured (D) states.

the 1970s by Privalov and co-workers [41]. The calorimetric criterion relates thermodynamic cooperativity to a bimodal distribution of the enthalpy, describing a conformational population peaked around the native and the unfolded states and practically zero at intermediate enthalpies, at the midpoint of the folding transition (Figure 5). More recently, it was suggested that a kinetic criterion known as the "chevron-plot" is the definitive hallmark of two-state folding cooperativity [42]. Why? The "chevron-plot" is a V-shaped graph, that results from plotting the logarithm of $k$, the relaxation rate constant, as a function of the denaturant concentration (at constant temperature), where one observes that the folding rate constant, $k_f$, dominates at low denaturant concentration while the unfolding rate constant, $k_u$, dominates at high denaturant concentration. The chevron-plot was shown to be a more restrictive criterion for cooperative folding behaviour than the calorimetric criterion in the sense that it is not present in all of the two-state proteins that passed the calorimetric cooperative test. In other words, thermodynamic cooperativity is a necessary but not a sufficient condition for kinetic cooperativity (see Ref. [42] for a recent review on cooperativity principles in protein folding). The folding of lattice polymers (Gō and others) appears to be relatively non-cooperative in this sense [42, 43] (note however that, in a broader sense as pointed out in the previous section, different degrees of cooperative behaviour may be observed in lattice models.). Based on these observations Kaya and Chan [44] suggested that if proteinlike cooperativity is the main drive for geometry-dependent kinetics, that appears to be lacking in simple lattice models, it is not surprising that their folding rates are weakly correlated with contact order. The question is then what happens in these models if proteinlike cooperativity is enhanced ? Kaya and Chan [44] addressed this question in the context of a new Gō-type interaction scheme based on

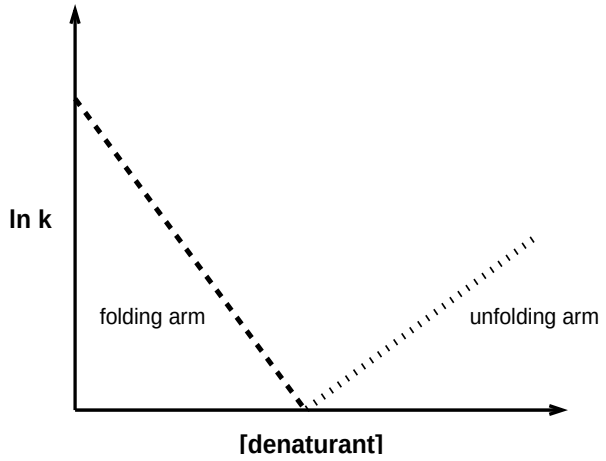


FIG. 6: A pictorial representation of a 'chevron-plot', a distinctive feature of experimental two-state folding kinetics. The relaxation rate constant, $k$, is the sum of the folding and unfolding rates constants, $k = k_f + k_u$. The folding rate constant dominates at low denaturant concentration while the unfolding rate constant dominates at high denaturant concentration.

the crucial assumption that a favourable correlation exists between local conformational preferences and nonlocal interactions, i.e., nonlocal interactions are stabilised when the chain segments around the native residues are in their native conformations (the model is suggested by the experimental observation that secondary structural elements are stable only in the native fold [45]). This correlation is most simply modelled as a (non-additive) many-body 'attraction' that promotes the folding of large portions of the native structure which in turn increases the energy gap between the native and unfolded conformations. Kaya and Chan have found that this generalized Gō model exhibits thermodynamic cooperativity and linear chevron plots similar to those observed experimentally for real two-state folders. Moreover the model yields folding rates that are logarithmically well correlated ($r = 0.94$) with the contact order parameter. The way in which these many-body correlations arise in general (or are 'encoded' in the primary sequence) remains an open question.

## VII. CONCLUSIONS

In the present work we have reviewed some results and concepts that emerged in the field of protein folding during the last few years, based on results obtained through Monte Carlo simulations of simple lattice models.

There is experimental evidence that the folding of many small, single domain, two-state proteins occurs in smooth energy landscapes and that the folding kinetics is correlated to the geometry of the native state. In this context Gō- and Gō-type models are becoming increasingly popular since lattice-polymers modelled by the Gō potential exhibit smooth energy landscapes where the effects of native geometry dominate. A recent finding obtained from simulations on these models is that the long-range contacts play a determinant role in the folding kinetics and that this effect is geometry-dependent.

However, recent results on the traditional Gō potential with additive pairwise interactions suggest that these models fail to capture the proteinlike cooperativity of real two-state folding kinetics and the question of wether realistic folding may be addressed within the scope of simple lattice models becomes relevant. In particular, the idea that a geometry-dependent kinetics may result from the type of cooperativity exhibited by real two-state folders has proven a new challenge for lattice models. The development of new interaction schemes based on non-additive many-body interactions leading to Gō-type models with protein-like cooperativity appears to be the next step towards a more realistic lattice modelling of two-state folding kinetics.

It is an honour and a great pleasure to contribute to this issue in honour of Ben Widom. Simple lattice models for complex fluids owe a lot to Ben's imagination and creativity. Thank you Ben for leading the way and Happy Birthday!

## VIII. ACKNOWLEDGMENTS

PFNF would like to thank Fundação para a Ciência e a Tecnologia for financial support through grant number BPD10083/2002.

* Electronic address: patnev@alf1.cii.fc.ul.pt

REFERENCES

[1] Anfinsen CB. Principles that govern the folding of protein chains, Science 1973;181:223-230

[2] Levinthal CJ. Are there pathways for protein folding?, Chim. Phys. 1968; 65:44-45

[3] Bryngelson J, Wolynes P. Spin glasses and the statistical mechanics of protein folding, Proc Natl Acad Sci USA 1987;84:7524-7528.

[4] Sali A, shakhnovich, Karplus, M. How does a protein fold? Nature 1994;369:248-251

[5] Bryngelson J, Wolynes P. Intermediates and barrier crossing in a random energy model (with applications to protein folding), J Chem Phys 1989;93:6902-6915

[6] Chan HS, Kaya H, Shimizu S. Computational Methods for Protein Folding: Scaling a Hierarchy of Complexities. Current Topics in Computational Molecular Biology 2002, Jiang T, Xu Y, Zhang MQ (Eds.), The MIT Press, Cambridge, Massachusetts, USA pp.403-447

[7] Abkevich, VI, Gutin, AM, and Shakhnovich, EI. Specific nucleus as the transition state for protein folding: Evidence from a lattice model, Biochemistry 1994;33:10026-10036

[8] Fersht, A. Optimization of rates of protein folding: the nucleation-condensation mechanism and its implications, Proc Natl Acad Sci USA 1995;92:10869-10873

[9] Fersht, A. Structure and Mechanism in Protein Science: Guide to Enzyme Catalysis and Protein Folding 1999, W. H. Freeman, New York; pp. 631

[10] Dinner A, Sali A, Smith LJ, Dobson CM, Karplus M. understanding protein folding via free-energy surfaces from theory and experiment. TIBS 2000;25:331-339

[11] Gillespie B., Plaxco KW. Two-state folding: theory and experiment. Annu Rev Biochem 2004;73:837-859

[12] Plaxco KW, Simmons KT, Baker D. Contact order, transition state placement and the refolding rates of single domain proteins. J Mol Biol 1998;277:985-994

[13] Plaxco KW, Simmons KT, Ruczinski I, Baker D. Topology, stability, sequence and length: Defining the determinants of two-state protein folding kinetics. Biochemistry 2000;39:11177:11183

[14] Gromiha MM, Selvaraj S. Comparison between long-range interactions and contact order in determining the folding rate of two-state folders: Application of long-range order to folding rate prediction. J Mol Biol 2001;310:27-32

[15] Zhou HY, Zhou YQ. Folding rate prediction using total contactdistance. Biophys J 2002;82:458-463

[16] Gō N, Taketomi H. Respective roles of short- and long-range interactions in protein folding. Proc Natl Acad Sci USA 1978;75:559-563

[17] Shakhnovich EI,Proteins with selected sequences fold into unique native conformation, Phys. Rev. Lett. 1994;72:3907?3910

[18] Miyazawa, S, Jernigan, R. Estimation of effective interresidue contact energies from protein crystal structures: quasi-chemical approximation. Macromolecules 1985;18:534-552.

[19] Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of state calculations by fast computing machines. J Chem Phys 1953;21:1087-1092

[20] Shakhnovich EI, Gutin AM, Engineering of stable and fast-folding sequences of model proteins, Proc Natl Acad Sci USA 1993;90 (1993):7195-7199.

[21] Gutin AM, Abkevich VI, Shakhnovich, EI. Chain lenght scaling of protein folding. Phys Rev Lett 1996;77:5433-5436

[22] Gutin A, Sali A, Abkevich V, Karplus M, Shakhnovich EI. Temperature dependence of the folding rate in a simple protein model: Search for a "glass" transition. J Chem Phys 1998:108;6466-6483

[23] Cieplak M, Hoang TX, Li MS. Scaling of folding properties in simple models of proteins. Phys Rev Lett 1999;83:1684-1687

[24] Faisca PFN, Ball RC. Thermodynamic control and dynamical regimes in protein folding. J Chem Phys 2002;116:7231-7238

[25] Myers JK, Oas TG. Mechanism of fast protein folding. Ann. Rev. Biochem. 2002;71:783-815.

[26] Jackson, S. How do small single-domain proteins fold? Fold. Des. 1998;3:81-91

[27] Dinner AR, Sali A, Karplus M. The folding mechanism of larger model proteins: Role of native structure, Proc Natl Acad Sci USA 1996;93:8356-8361

[28] Thirumalai, D. From Minimal Models to Real Proteins: Time Scales for Protein Folding Kinetics. J. Phys. I (France) 1995;5:1457-1467

[29] Galzitskaya, OV, Garbuzynskiy, SO, Ivankov, DN, Finkelstein, AV. Chain Length is the Main Determinant of the Folding Rate for Proteins with Three-State Folding Kinetics, Proteins 2003;51:162-166

[30] Abkevich VI, Gutin AM, Shakhnovich EI. Impact of local and non-local interactions on thermodynamics and kinetics of protein folding. J Mol Biol 1995;252:460-471

[31] Doyle R, Simons K, Qian H, Baker D. Local interactions and the optimization of protein folding. Proteins 1997;29:282-291

[32] Gromiha MM and Selvaraj S. Importance of long-range interactions in protein folding. Biophys Chem 1997;77:49-68

[33] Unger E, Moult J. Local interactions dominate folding in a simple protein model. J Mol Biol 1996;259:988-994

[34] Govindarajan S, Goldstein RA. Optimal local propensities for model proteins. Proteins 1995;22:413-418

[35] Faisca PFN, Telo da Gama MM, Nunes A. The Gō model revisited: Native structure and the geometric coupling between local and long-range contacts, Proteins 2005 (accepted), q-bio.BM/0411044.

[36] Faisca PFN, Ball RC. Topological complexity, contact order and protein folding rates, J Chem Phys 2002;117:8587-8591

[37] Gillespie B, Plaxco KW. Nonglassy kinetics in the folding of a simple single-domain protein. Proc Natl Acad Sci USA 2000;97:12014-12019

[38] Shea JE, Onuchic JN, Brooks CL. Probing the folding free energy landscape of the src-SH3 protein domain. Proc Natl Acad Sci USA 2002;99:16064-16068

[39] Jewett AI, Pande VS, Plaxco KW. Cooperativity, smooth energy landscapes and the origins of topology-dependent protein folding rates. J Mol Biol 2003;326:247-253

[40] Faisca PFN, Telo da Gama MM, Ball RC. Folding and form: Insights from lattice simulations. Phys Rev E 2004;69:051917

[41] Pfeil, W, Privalov PL. Thermodynamic investigations of proteins. I. Standard functions for proteins with lysozyme as an example. Biophys Chem 1976;4:23-32

[42] Chan HS, Shimizu S, Kaya H. Cooperativity Principles in Protein Folding. Methods in Enzymology 2004;380:350-379

[43] Kaya H, Chan HS. Origins of Chevron Rollovers in Non-Two-State Protein Folding Kinetics. Phys Rev Lett 2003;90:258104

[44] Kaya H, Chan HS, Contact order dependent protein folding rates: Kinetics consequences of a cooperative interplay between favourable nonlocal interactions and local conformational preferences. Proteins 2003;52:524:533

[45] Dill, KA. Dominant forces in protein folding. Biochemistry 1990;29:7133-7155